Firm Reputation and Horizontal Integration*

Hongbin Cai† Ichiro Obara‡

March 14, 2008.

Abstract

We study effects of horizontal integration on firm reputation. In an environment where customers observe only imperfect signals about firms' effort/quality choices, firms cannot maintain good reputation and earn quality premium forever. Even when firms choose high quality, there is always a possibility that a bad signal is observed. Thus, firms must give up their quality premium, at least temporarily, as punishment. A firm's integration decision is based on the extent to which integration attenuates this necessary cost of maintaining a good reputation. Horizontal integration leads to a larger market base for the merged firm, which leads to a more effective punishment and a better monitoring by eliminating idiosyncratic shocks in many markets. But it also allows the merged firm to deviate in a more sophisticated way: the merged firm may deviate only in a subset of markets and pretend that a bad outcome in those markets is observed by accident. This negative effect becomes very severe when the size of the merged firm gets larger and there is non-idiosyncratic firm-specific noise in the signal. These effects give rise to a reputation-based theory of the optimal firm size. We show that the optimal firm size is smaller when (1) trades are more frequent and information is disseminated more rapidly; or (2) the deviation gain is smaller compared to the quality premium; or (3) customer information about firms' quality choices is more precise.

Keywords: Reputation; Integration; Imperfect Monitoring; Theory of the Firm; Merger

JEL Classification: C70; D80; L14

*We thank seminar participants at Brown University, Stanford University, UC Berkeley, UCLA, UCSB, UC Riverside, UIUC and USC for helpful comments. All remaining errors are our own.

†Guanghua School of Management and IEPR, Peking University, Beijing, China 100871. Tel: (86) 10-62765132. Fax: (86) 10-62751470. E-mail: hbcai@gsm.pku.edu.cn

‡Department of Economics, UCLA, 405 Hilgard Ave, Los Angeles, CA 90095-1477. Tel: 310-794-7098. Fax: 310-825-9528. E-mail: iobara@econ.ucla.edu

1 Introduction

Reputation has long been considered critical for firm survival and success in the business world. Since the seminal work of Kreps (1990), the idea of firms as bearers of reputation has become increasingly important in the modern development of the theory of the firm. For example, Tadelis (1999, 2002), Mailath and Samuelson (2001), and Marvel and Ye (2004) develop models of firm reputation as tradable assets and study the market equilibrium for such reputation assets. Klein and Leffler (1981) and Hörner (2002) analyze how competition helps firms build good reputations when their behavior is not perfectly monitored by customers. These studies provide very useful insights into how firm reputation can be built, maintained and traded. However, for reputation to be a defining feature in the theory of the firm, an important question needs to be answered: How does firm reputation affect the boundaries of the firm?[1]

In this paper we build a simple model to study the effects of horizontal integration on firm reputation. We consider an environment where firms produce experience goods in the sense that customers cannot observe product quality at the time of purchase, but their consumption experience provides noisy public information about product quality (e.g., consumer ratings).[2] Absent proper incentives, firms will tend to shirk on quality to save costs, making customers reluctant to purchase. Using a model of repeated games with imperfect monitoring, it is easy to show that as long as firms care sufficiently about the future, they can establish reputations of high quality and earn quality premium while building customers loyalty.[3] However, unlike the case with perfect monitoring, firm reputation can be sustained only if the public signal about a firm's choices is above a certain cut-off point in every period. With positive probability the public signal will fall below the cut-off point, in which case firm reputation will be lost: either customers will never buy again or the firm must pay large financial penalties to win back previous customers.

We then consider a situation where several firms, each serving an independent and symmetric market, merge into one large firm.[4] Horizontal integration leads to a larger market base for the

[1] The boundary of the firm question was first raised by the classical work of Coase (1937). Several influential theories have been proposed to answer the question, for example, Alchian and Demsetz (1972), Williamson (1985), and Hart (1995). Holmstrom and Roberts (1998) offer a review and critique of these theories.

[2] Professional services, food services, and consumer durable goods are standard examples of experience goods.

[3] Our analysis is an application of the theory of repeated games with imperfect monitoring, see, e.g., Green and Porter (1984), Abreu, Pearce, and Stacchetti (1986, 1990), and many others.

[4] Note that we consider horizontal integration of firms that produce similar products (the assumption of symmetric

merged firm, which allows a more effective punishment and better monitoring by eliminating all idiosyncratic shocks across the markets. On the other hand, horizontal integration gives the merged firm more room for sophisticated deviations; it may deviate only in a subset of markets and pretend that a bad outcome in those markets was observed by an accident. This negative effect becomes more severe as the firm size gets larger when there is non-idiosyncratic firm-specific noise in the perfect public signal. These two effects on reputation building give rise to meaningful trade-offs for horizontal integration, leading to a reputation-based theory of the optimal firm size. We characterize the optimal level of integration and provide a clear comparative statics result regarding the optimal size of the firm. We show that the optimal size of the firm is smaller (or, non-integration is more likely to dominate integration) when (1) trades are more frequent and information is disseminated more rapidly; or (2) the deviation gain is relatively smaller compared to the reputation premium; or (3) customers information about firms' choices is more precise. We also provide sufficient conditions under which non-integration is optimal.

The results of our paper shed light on the patterns of horizontal integration observed in the real world. For example, horizontal integration such as franchising is very common in industries that mainly provide services to travelers, e.g., hotels and car rentals.[5] In these industries, customers interact with a firm infrequently and the customer base of a firm tends to be quite heterogenous, which corresponds to more discounting (smaller discount factors) and larger communication noises in our model. As our results show, in such cases independent firms cannot build reputation effectively, and horizontal integration can improve on reputation building. Similarly, in industries providing services to both travelers and locals such as taxicabs and convenience stores, horizontal integration (either as franchising like the Seven-Eleven stores, or mergers of taxicab companies) seems to be quite common, though less as common as in purely travel industries.

For another example, chains are more common in the fast food sector than they are among high end restaurants. Fast food restaurants provide more homogenous products than high end

markets). Our model does not directly apply to firms with multiple product lines that are obviously of different quality levels, e.g., Holiday Inn and Holiday Inn Express, or Toyota's wide range of models, from Tercel to Lexus. Such cases (brand expansion) are analyzed by Andersson (2002) from reputation concerns.

[5] Note that we focus on the reputation pooling aspect of franchising in the paper, and ignore the ownership and incentives issues of franchising that has been analyzed extensively in the existing literature. Two essential features common in many franchises, trademark (brand names carrying collective reputation) and quality control, correspond nicely to our model.

restaurants, thus their profit margins are on average smaller than high end restaurants. Our results suggest that if payoffs from maintaining a good reputation are greater relative to deviation gains (e.g. high profit margin restaurants), non-integration is more likely to dominate integration.[6] While other explanations are certainly possible in these examples, our theory provides a new perspective and offers a new insight into horizontal integration that potentially can be tested with real world data. In fact, in a recent empirical paper that analyzes reputation incentives for restaurant hygiene, Jin and Leslie (2004) find evidence consistent with our theory. For example, they find that restaurant chains are more likely to be found in tourist locations. Moreover, "regions where independent restaurants tend to have relatively good quality hygiene, the incremental effect on hygiene from chain affiliation is lower."

To understand the basic ideas, consider several independent firms serving separate markets and suppose that they merge into one single firm. The integrated firm makes effort/quality decisions in the production process and allocates products to all the markets it serves. Customers in each market observe some noisy signal about the firm's product quality in all the markets. We first demonstrate that in the best equilibrium, only the average effort or the average signal matters. Specifically, as long as the average signal is above a certain cut-off point, the firm provides high quality goods in all the markets; otherwise either customers in all the markets desert the firm forever or the firm pays large financial penalties to the customers in all markets.[7] This cut-off point determines how long the firm's reputation will be sustained, thus the level of the optimal reputation equilibrium payoff. The lower the cutoff point is, the more efficient the optimal reputation equilibrium is. Then the optimal firm size of a firm is the size for which the lowest cut-off point is obtained so that reputation lasts the longest, thus the expected profit per market is maximized.

In our model, integration has three effects on reputation: a positive size effect, a positive

[6] The pattern of horizontal integration in the food industry is also consistent with the previous point: compared to fast food restaurants (especially those along highways or in airports), high end restaurants are more focused on serving local communities. Thus, less discounting (larger discount factors) and smaller communication noises make non-integration more attractive for high end restaurants.

[7] This is consistent with the observation that customers usually care about public signals about a firm's aggregate choices or overall performance such as its product quality ranking and rating of consumer satisfaction. Public signals about each branch's choice may not be available or too noisy to be useful. For example, it can prove very difficult to discern accounting records for each of the firm's divisions since there are numerous ways to allocate costs and revenues within the firm. But even in cases where public signals of product quality are available in each market, our result (Lemma 3) suggests that it is sufficient to look at the aggregate signal about the firm's overall quality.

information effect, a negative deviation effect. First, a large firm size helps reputation building by making more severe punishment possible (e.g., shutting down the whole firm in every market) for a fixed magnitude of deviation (e.g., choosing low quality in only one market). The merged firm has more to lose, thus has more incentives to maintain reputation for a given cutoff point. This effect makes it possible to use a lower the cutoff point to make reputation more long lasting.[8]

The second effect of integration on reputation, the information effect, is that integration may allow information aggregation across markets and thus make it easier for customers to monitor a firm serving a large number of markets. Since an integrated firm's reputation is contingent on the public signal averaged over the markets it serves, its reputation mechanism depends on the informativeness of the average signal. Since market-specific idiosyncratic shocks are assumed to be independent, they are washed away by the law of large numbers as the size of the firm becomes larger. Thus reputation building becomes relatively easier for large firms.

The last effect of integration on reputation is that the merged firm has more opportunities for deviation than independent firms. A firm that serves n markets can deviate in any $m \leq n$ markets, and thus has to satisfy n incentive constraints to maintain its reputation. This, of course, impedes reputation-building. This constraint is especially strong when the large firm deviates only in a tiny fraction of markets because bad outcomes in a few markets are indistinguishable from noise. Indeed we can show that under mild conditions, the single-market deviation constraint is the only binding incentive constraint. While market-specific idiosyncratic shocks are washed away as the firm size increases, the firm-specific technological noise do not disappear. Since the size of noise in the average signal is approximately constant and the size of one-market deviation gets smaller (at the rate of $\frac{1}{n}$) as $n \to \infty$, it becomes more difficult to prevent the one-market deviation for a larger firm.

These three effects of horizontal integration on reputation present meaningful trade-offs regarding firm size When the positive size effect and information aggregation effect dominate the negative deviation effect, then we expect firms to optimally choose a greater degree of horizontal

[8] To be more precise, it is not firm size per se that matters. If an independent firm expands so that its payoffs in all contingencies simply scale up, its incentives to build reputation will not be affected at all. When two independent firms merge into one, what is important is that the merged firm makes joint decisions for both branches and its customers understand this. Hence if it appears that the firm has cheated somewhere, all its customers everywhere will punish it by desertion. This idea is first shown in Bernheim and Whinston (1990) and appears in subsequent papers such as Matsushima (2001) and Andersson (2002).

integration. Otherwise, the optimal size of the firm tends to be smaller, and even non-integration may be optimal. We show that the optimal size of a firm is bounded under reasonable conditions. Typically the negative deviation effect eventually dominates the other two effects as a firm becomes larger and larger. As a result, the optimal firm size can be obtained in our framework. This is in sharp contrast with the existing literature on reputation building. Although based on the model of repeated games like our paper, this literature emphasizes only positive aspects of integration as mentioned below, thus does not provide a theory of the optimal level of integration.

Our paper is closely related to the literature on multimarket contacts, e.g., Bernheim and Whinston (1990) and Matsushima (2001). Bernheim and Whinston (1990) show that in the perfect monitoring setting, two firms may find it easier to collude if they interact in multiple markets in which they have uneven competitive positions than if they interact in a single market. Matsushima (2001) considers the setting of imperfect monitoring and proves that two firms can approach perfect collusion when the number of market contacts goes to infinity. In these papers, merger always dominates independence because each market is completely independent from the other markets. Thus these models are not suited to analyze the bound of firm size. Our paper demonstrates that a bound on the firm size may naturally arise when there is a firm-specific production noise which has some common component across the markets served by the same firm.

In terms of motivations, our paper is perhaps most closely related to Andersson (2002), Gutman and Yekouel (2002) and Fishman (2005), all of which study the effects of integration on firm reputation in models with perfect monitoring. In Andersson (2002), a firm producing multiple products may increase its total profits (relative to independent firms producing those products), because pooling the incentive constraints in the multiple markets may allow the firm to increase its prices.[9] In Gutman and Yekouel (2002) and Fishman (2005), integration facilitates reputation formation by increasing the number of consumers each firm serves, which increases the chance that new consumers learn about the firms' performance from pervious consumers in their settings.[10] In our model, firm choices are imperfectly monitored, and we consider integration across symmetric

[9]The applications Andersson (2002) considers are brand extensions or "umbrella branding" whereby a firm produces different kinds of products under one brand (e.g., Porsche watches). For recent contributions and a summary of the literature, see Cabral (2000).

[10]In their models, each consumer can tell J consumers of the next generation about a firm's performance, and then each of them can pass on the information to J consumers of the next next generation, and so on. Thus, monitoring is not exactly perfect but becomes perfect over time.

markets.[11] In contrast to our model, a firm's profit can only increase monotonically in its size in all the papers mentioned above (e.g. the bigger, the better).[12]

The rest of the paper is organized as follows. The next section presents the model. Then in Sections 3 and 4, we characterize the best reputation equilibrium of the game for the firm under non-integration and integration, respectively. Comparing the equilibrium outcomes in the two cases, in Section 5 we obtain the main results about the optimal firm size and examine how it is affected by the parameters of the model. Concluding remarks are in Section 6.

2 The Model

There are a large number of separate markets, in each of which a long-lived firm sells its products to its customers.[13] Time is discrete and the horizon is infinite. Customers in each market are identical, and the firms and their respective markets are symmetric. In each period, the firm in each market and its customers play the following stage game. At the beginning of a period, the firm, who we assume has price-setting power, sets the price p for the period.[14] Then the firm and the customers play the following game. The customers decide whether to purchase one unit of the firm's products. If they do not buy from the firm, both the customers and the firm get a payoff of zero. If they decide to buy from the firm, their payoffs depend on the firm's product quality. The firm decides whether to exert high effort e_h (or, provide high quality) or exert low effort e_l (or, provide low quality), where e_h and e_l are both real numbers and $e_h > e_l$. The firm incurs an effort/quality cost of c_h (c_l) for providing high (low) quality, where $c_h > c_l$. The customers' expected benefit is v_h if the firm chooses e_h and is v_l if the firm chooses e_l, where $v_h > v_l$. Given p, the stage game is depicted below in the normal form. Equivalently one can think of an extensive

[11] In both Bernheim and Whinston (1990) and Andersson (2002), firm size will have no effect if all markets are symmetric. Unlike Gutman and Yekouel (2002) and Fishman (2005), the ratio of consumers to firms does not play a learning role in our model.

[12] Fishman and Rob (2002) study a model of investment in reputation in which firms' product qualities are perfectly observed by some customers, and show that bigger and older firms have better reputations. Rob and Sekiguchi (2004) analyze reputation formation under imperfect monitoring in a repeated duopoly setting.

[13] It is not difficult to introduce competition across the markets. We chose not to do so just because we like to focus on the reputation-building effect of integrations rather than the typical anti-competitive effect of integrations.

[14] Our analysis and the main results of the paper will not be affected significantly if the firm does not have full price-setting power.

form game in which the customers move first with their purchase decisions.

		Firm: Low	Firm: High
Customers	*Don't Buy*	0 , 0	0 , 0
	Buy	$v_l - p$, $p - c_l$	$v_h - p$, $p - c_h$

We assume $v_h - c_h > 0 > v_l - c_l$: high effort/quality is more efficient than no trade, which in turn is more efficient than low effort/quality. Since $c_h > c_l$, e_l (weakly) dominates e_h for the firm. Hence, for any price $p > v_l$, the unique equilibrium outcome is $(Don't\ Buy, Low)$, resulting in payoffs $(0, 0)$. The outcome $(Buy, High)$ is the first best efficient in terms of total surplus and Pareto-dominates $(Don't\ Buy, Low)$ for $p \in (c_h, v_h)$. However, this efficient outcome is not attainable without reputation effects. Our stage game is in the spirit of Kreps (1990), who highlights the firm's incentive problem in a one-sided Prisoners' Dilemma game.

We suppose that in each market there are a large number of identical customers that are anonymous to the firm in the market. Since an individual customer's behavior is not observable by the firm, customers will maximize their current period payoffs. Alternatively, we can assume customers purchase the products only once (i.e., short-lived customers), in which case they also maximize current period payoffs.[15]

If the firm's effort in each period were publicly observable, it would be straightforward to show that the efficient outcome $(Buy,\ High)$ can be supported when future is sufficiently important to the firm. Let δ be the firm's discount factor. It can be easily checked that for any price $p \in (c_h, v_h]$, the first best outcome is attainable in every period if and only if $\delta \geq (c_h - c_l)/(p - c_l)$. To maximize its profit, the firm will set price $p = v_h$.

In many cases, however, the firm's effort cannot be perfectly observable, especially for experience goods. For example, given the firm's effort, there are unavoidable uncertainties (e.g., machine malfunctioning, human errors) in production processes that introduce random shocks into product qualities. When customers purchase the products just once (i.e., short-lived), experiences of the current period customers may be communicated to future customers with substantial noise (e.g., consumer on-line ranking/comments). In such cases, the firm's past effort/quality choices can only be imperfectly observed by the customers in the future.

[15] Our assumption that customers maximize current period payoffs implies that the folk theorem result of Fudenberg, Levine and Maskin (1994) does not apply.

Given these observations, we consider an environment in which a firm's effort is not public information, but rather its noisy public signal $y \in \Re$ becomes available at the end of each period in each market. Suppose that this firm is serving market j. Given the firm's technology, its product quality in market j is given by $q_j = e_j + \eta$, where $\eta \sim N(0, \sigma_\eta^2)$ is a firm-specific production noise and is independent across periods. Since firms' technology choices are made independently, production noise η is assumed to be independent across markets as well. In addition to the production noise, the public signal also contains a market-specific demand noise component $\theta_j \sim N(0, \sigma_\theta^2)$, which is independent across markets and across periods, and independent from production noise. Thus, the public signal in market j is $y_j = e_j + \eta + \theta_j = e_j + \epsilon_j$, where $\epsilon_j \sim N(0, \sigma^2 = \sigma_\eta^2 + \sigma_\theta^2)$ represents the total noise. This information structure is quite natural, and is commonly used in the existing literature.

We should note, however, that many of our results can be extended to a more general information structure, where the signal is drawn from a distribution function $F(y|e)$ with a positive density function $f(y|e)$ that satisfies the strict Monotone Likelihood Ratio Property in the sense of Milgrom (1981). An earlier version of this paper, Cai and Obara (2004), uses this general formulation. When presenting our results, we shall point out whether they can be extended to more general information structure.

Following Green and Porter (1984) and Fudenberg, Levine and Maskin (1994), we focus on (pure strategy) perfect public equilibria of the game. In a perfect public equilibrium, players' strategies depend only on the past realizations of the public signals. For periods $t = 2, 3, ...$, the public history h_t is the sequence of signal realizations and prices in period $t-1$ and before. The customers will base their period t decisions on (h_t, p_t). The firm's pricing decision in period t depends on h_t and its effort/quality decision in period t on (h_t, p_t). In equilibrium, given its full price setting power, the firm will always set its price equal to either the customers' expected benefit from consuming its products or its production cost, whichever is larger.

We will characterize the perfect public equilibria of the game that yield the greatest average payoff for the firm, first for the non-integration case in which firms are independent, and then for the integration case, in which firms merge into one big firm which serves multiple markets. Since firms make decisions about integration or disintegration to maximize their value, by comparing the best equilibrium outcomes in the non-integration and integration cases, we derive conditions under which integration is better than non-integration or vice versa. We simply call a perfect public equilibrium that yields the greatest average payoff for the firm a "best equilibrium".

3 Best Equilibrium for the Non-Integration Case

We start with the non-integration case. In the non-integration case, firms are independent decision makers, so the public signal in one market will not affect the other markets at all, even if it is observable to the participants in the other markets. Since firms and markets are symmetric, we focus on a representative firm and its market (thus dropping subscript j).

As is typical in repeated games, there can be many perfect public equilibria in our game, many of which can involve complicated path-dependent strategies. However, it turns out that the best equilibria in our game have a very simple structure. Define a *cut-off trigger strategy equilibrium* as follows: the firm and its customers play $(Buy,\ High)$ in the first period and continue to choose $(Buy,\ High)$ as long as y stays above some threshold $\widetilde{y}$, and play the stage game Nash equilibrium $(Don't\ Buy,\ Low)$ forever once y falls below the threshold $\widetilde{y}$. The following lemma, shows that the best equilibrium must be a cut-off trigger strategy equilibrium whenever it is a nontrivial one.[16] This result holds for any general distribution $F(y|e)$ satisfying the Monotone Likelihood Ratio Property.

Lemma 1 *The best equilibrium for the firm is either a cut-off trigger strategy equilibrium with $p = v_h$ in every period, or the repetition of the stage game Nash equilibrium $(Don't\ Buy,\ Low)$.*

Proof: See the Appendix.

Note that in our game, the firm's pricing decision can be treated separately from its quality decision and customers' purchase decision. Since customers maximize their current period payoffs, they will purchase if and only if the price is not greater than their expected valuation (i.e., v_h or v_l, depending on their expectation of the firm's quality choice). Thus, in a best equilibrium for the firm, it can charge a price that equals to the customers' expected benefits. When the firm's reputation is good and is expected to provide high effort in the current period, it sets price $p = v_h$. If the firm loses its reputation and is expected to choose low effort, the highest price acceptable to customers is v_l, which is not sufficient to cover c_l by our assumption. Therefore, whenever customers expect the firm to choose low effort, the equilibrium outcome is no trade and price is trivially indeterminate. Since the firm's optimal pricing decision is straightforward, we focus our analysis on its quality decisions and reputation building.

[16]This is similar to Theorem 7 of Abreu, Pearce, Stacchetti (1990), which proves the necessity of bang-bang continuation payoffs for optimal equilibria.

Let us fix some terminology and notation. We will sometimes call a *stationary cut-off trigger strategy equilibrium* "reputation equilibrium". The periods in which the firm's reputation is good and can thus earn quality premium are called the "reputation phase"; otherwise they are called the "punishment phase". Let π be the firm's expected profit averaged over all periods in a best equilibrium. Define $r = p - c_h \, (= v_h - c_h)$ to be the firm's current period payoff if it exerts high effort ("honesty payoff"), and $d = c_h - c_l$ to be the cost differential of high and low efforts. If the firm chooses low effort, its current period payoff is $p - c_l = r + d$, so d is the firm's gain from deviation.

Let $\tilde{y}$ be the cut-off signal used in the equilibrium. Since the public signal is given by $y = e + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$, the probability of reputation continuing conditional on effort e is $1 - F(\tilde{y}|e) = 1 - \Phi\left(\frac{\tilde{y}-e}{\sigma}\right)$, where Φ is the standard normal distribution function. Then the firm's average payoff in the equilibrium, π, must satisfy the following value recursive equation:

$$\pi = (1-\delta)r + \delta(1 - F(\tilde{y}|e_h))\pi = (1-\delta)r + \delta\left(1 - \Phi\left(\frac{\tilde{y} - e_h}{\sigma}\right)\right)\pi \tag{1}$$

Equation (1) says that the firm's per period value in the equilibrium is the sum of its current period profit averaged out over time, $(1-\delta)r$, plus the expected average value from continuation, $\delta(1 - F(\tilde{y}|e_h))\pi$.[17]

For the firm to be willing to choose e_h, the incentive compatibility constraint requires

$$\pi \geq (1-\delta)(r+d) + \delta(1 - F(\tilde{y}|e_l))\pi = (1-\delta)(r+d) + \delta\left(1 - \Phi\left(\frac{\tilde{y} - e_l}{\sigma}\right)\right)\pi \tag{2}$$

The right hand side of Equation (2) if the firm's average payoff from choosing e_l, which consists of the current period profit averaged out over time, $(1-\delta)(r+d)$, plus the expected average value from continuation, $\delta(1 - F(\tilde{y}|e_l))\pi$.

Any pair of $(\pi, \tilde{y})$ that satisfies both Equations (1) and (2) gives rise to an equilibrium in which the firm will choose high effort every period and customers continue to purchase as long as $y \geq \tilde{y}$.

Lemma 2 *The IC constraint of Equation (2) must be binding in the best reputation equilibrium.*

[17] Following the convention of the repeated game literature, we measure a firm's payoff as its expected profit averaged over the infinite horizon, instead of its total discounted expected payoff. These two measures differ by a factor of $1 - \delta$, but the former slightly simplifies notation.

Proof: Consider any cut-off trigger strategy equilibrium with $(\pi, \tilde{y})$ such that Equation (2) holds as a strict inequality. Then we can decrease $\tilde{y}$ without affecting the IC constraint. However, as is clear from Equation (1), this reduces the value of $F(\tilde{y}|e_h)$, thus increases π. Contradiction. *Q.E.D.*

By Lemma 2, we can solve for the cut-off point $\tilde{y}$ and the firm's average profit π in the best reputation equilibrium (if it exists) from Equation (1) and Equation (2) as an equality. After some manipulation of terms we obtain

$$(1-\delta)d = \delta\left[F(\tilde{y}|e_l) - F(\tilde{y}|e_h)\right]\pi = \delta\left[\Phi\left(\frac{\tilde{y}-e_l}{\sigma}\right) - \Phi\left(\frac{\tilde{y}-e_h}{\sigma}\right)\right]\pi \tag{3}$$

This equation simply says that the current period gain from deviation averaged out over time (the LHS) equals the expected loss of future profit from deviation (the RHS).

It is convenient to focus on the normalized signal $k = \frac{y-e_h}{\sigma}$ instead of the signal y. Abusing notation slightly, we shall call k the public signal. Let $\tau = d/r$ be the ratio of the deviation gain to the honesty payoff, and $\triangle = e_h - e_l$ be the effort differential. Using Equation (1) to eliminate π from Equation (3), we obtain the following "*fundamental equation*:"

$$G(k) \equiv \frac{\Phi(k+\frac{\triangle}{\sigma}) - \Phi(k)}{\tau} - \Phi(k) = \frac{1-\delta}{\delta} \tag{4}$$

If there is a solution $\tilde{k}$ to the fundamental equation (4), then from Equation (1),

$$\pi = \frac{(1-\delta)r}{1-\delta[1-\Phi(\tilde{k})]} \tag{5}$$

Clearly π is a decreasing function of $\tilde{k}$. Hence the smallest solution to Equation (4) constitutes the cut-off (normalized) signal in the best reputation equilibrium. Thus we have

Proposition 1 *There exists a reputation equilibrium if and only if the fundamental equation (4) has a solution. If that is the case, then the smallest solution is the cut-off point in the best equilibrium. The firm's value in the best equilibrium is given by (5).*

It can be easily verified that Proposition 1 holds under any signal structure with a general distribution $F(y|e)$ replacing Φ in Equations (4) and (5).[18] By Proposition 1 and Lemma 1, the

[18] Lemma 2 is completely general. The uniqueness of the best cutoff point follows from the MLRP of $F(y|e)$.

existence of reputation equilibria hinges on whether Equation (4) has a solution. It is called the fundamental equation because its smallest solution determines the best cut-off point, which in turn determines the best equilibrium payoff for the firm through Equation (5). The characterization of the firm's average expected profit in Proposition 1 resembles that of Abreu, Milgrom and Pearce (1991), who study symmetric perfect public equilibria in repeated partnership games. It can be verified that Equation (5) is equivalent to

$$\pi = r - \frac{d}{\Phi(\tilde{k} + \frac{\triangle}{\sigma})/\Phi(\tilde{k}) - 1}$$

As in their model, here the firm's value equals its honesty payoff r minus an incentive cost (the second term of the RHS) that depends on the deviation gain d and the likelihood ration $\Phi(\tilde{k} + \frac{\triangle}{\sigma})/\Phi(\tilde{k})$, which measures how easily the public signal can reveal deviations.

It can be verified that the function $G(k)$ defined in Equation (4) is maximized at

$$k^* = -\frac{\triangle}{2\sigma} - \frac{\sigma \ln(1+\tau)}{\triangle}$$

Let δ^* be the discount factor that satisfies $\frac{1-\delta^*}{\delta^*} = G(k^*)$. In the Appendix, we show that the function $G(k)$ has the shape as shown in Figure 1.

Clearly Equation (4) has either no solution or two solutions, depending on whether δ is above or below δ^*.[19] When there are two solutions, the smaller solution $\tilde{k}$ is the cut-off point in the best equilibrium. Thus we have the following result:

Proposition 2 *There exists a reputation equilibrium if and only if $\delta \geq \delta^*$. When $\delta > \delta^*$, the cut-off point $\tilde{k}$ for the best equilibrium is decreasing in δ and $\triangle$, and increasing in τ and σ; the firm's average payoff is increasing in r, δ and $\triangle$, and decreasing in d and σ.*

Proof: See the Appendix.

Proposition 2 says that as long as the firm cares sufficiently about the future, reputation can be built in equilibrium in our model of imperfect monitoring. However, compared with the case of perfect monitoring (observable effort choices), reputation works less well. Reputation can break down with a positive probability (indeed almost surely in the long run) even on the equilibrium

[19] In a degenerate case, it has one solution for one particular δ^*.

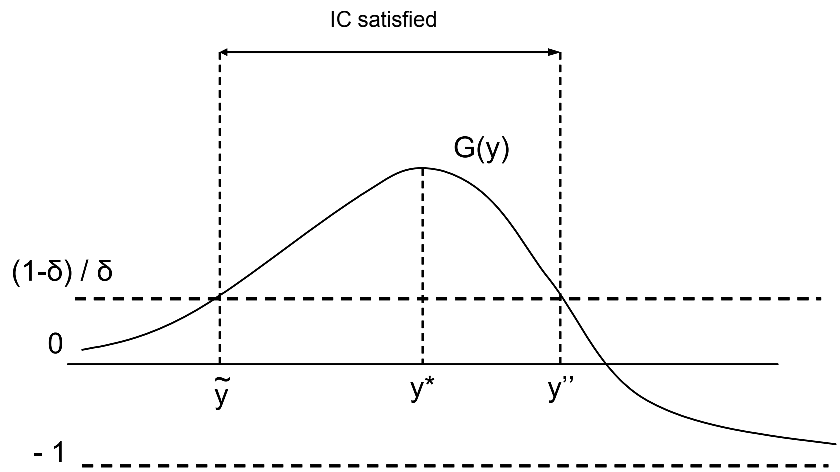

Figure 1: Graphic Illustration of the Fundamental Equation

path as in Green and Porter (1984). This is necessary to give the firm an incentive to stick to good behavior. In the case of perfect monitoring, no actual punishment is incurred in motivating the firm to choose high effort, since any deviation is perfectly detected.

Proposition 2 also establishes the comparative statics for the best equilibrium, which are all intuitive. It says that reputation will more likely be sustained if (i) the firm cares more about the future (greater δ); (ii) the public signal is more revealing about the firm's effort choice (greater $\triangle$); (iii) the gain from deviation in relative terms is smaller (smaller τ), or (iv) the public signal is less noisy (smaller σ). Note that in Equation (5), a smaller cut-off point leads to a higher average payoff (with r kept constant). All comparative statics follow from this simple observation. Except for the comparative statics about σ, the rest of Proposition 2 can be extended to the general distribution $F(y|e)$ under a technical assumption. [20]

So far we have assumed that the stage Nash equilibrium $(Don't\ Buy, Low)$ is played forever

[20] Specifically, the assumption is $\lim_{y\to-\infty} f(y|e_l)/f(y|e_h) > 1+\tau$. This assumption means that small y is sufficiently informative about a deviation to low effort. If this inequality is not satisfied, then G becomes a negative, always decreasing function, thus the fundamental equation has no solution so the only equilibrium of the game is the repetition of the stage Nash equilibrium ($Don't\ Buy$, Low). This condition is easily satisfied by the normal distribution with e_i being the mean, as the likelihood ratio goes to infinity as $y \to -\infty$.

in the punishment stage. However, both the firm and its customers have strong incentives to renegotiate and continue with their relationship when the signal falls below the cut-off point (especially considering that the firm did not do anything wrong on the equilibrium path). In other words, the above equilibrium is not renegotiation-proof. In Appendix B, we show that the firm's minmax payoff of 0 may be implemented in another equilibrium that is renegotiation-proof, in which the firm offers a large discount to the customers by drastically cutting its price when the signal is below the cut-off point. Thus, the same equilibrium outcome derived in this section can be supported in a renegotiation-proof equilibrium.

4 Best Equilibrium for the Integration Case

Now we analyze the integration case in which n firms merge into one big firm. We assume that the integrated firm adopts a common technology for all n markets (branches). Since the firm adopts a common technology, all its branches share the same production noise η.[21] This seems consistent with the observation that integrated firms often have centralized quality controls and try to maintain quality standards across markets and divisions. For example, franchised firms and chains typically have centralized supply systems and closely monitor branches and stores for quality controls.[22]

Once the big firm has chosen its technology, the big firm first chooses a price vector $(p_1, p_2, ..., p_n)$ and then an effort vector of $(e_1, e_2, ..., e_n)$ in the beginning of each period. The public signal in market $j$is given by $y_j = e_j + \eta + \theta_j$, where the common production noise η follows $N(0, \sigma_\eta^2)$ and the idiosyncratic market demand noise θ_j follows $N(0, \sigma_\theta^2)$. Hence y_j can be interpreted as a noisy signal of quality $q_j = e_j + \eta$ in jth market. Let $\epsilon_j = \eta + \theta_j$ be the total noise in market j., which is a normal random variable with mean 0 and variance $\sigma_\eta^2 + \sigma_\theta^2$.

A crucial assumption we make is that *all the customers of the integrated firm use a monitoring strategy based on the average signal* $y_n = \bar{e}_n + \bar{\epsilon}_n$, where $\bar{e}_n = \sum_1^n e_j/n$ and $\bar{\epsilon}_n$ is the average

[21] What we really need here is that technology shocks across the markets served by one big firm have *some* common components. In particular, our analysis extends to the case in which the production noise has some idiosyncratic shocks across markets; $\eta_j = \eta_0 + \xi_j$, where ξ_j are i.i.d. mean-zero random variables.

[22] An industry expert, Mark Siebert, writes that "Top franchisors know that brand maintenance means more than just marketing. It also means quality control. The best franchisors typically have field support personnel whose responsibility is to visit franchisees in the field and determine if they're living up to brand standards. .... Beyond field support, the best franchisors are huge advocates of training. " (http://www.entrepreneur.com/franchises/)

noise. Given this assumption, it is straightforward to extend Proposition 1 to the integration case, that is, we can focus on equilibria in which customers continue to buy if and only if the average signal is above some cutoff point $\widetilde{y}_n$.

Why customers may use only the average signal

There are two justifications for this assumption. First, customers of an integrated firm often pay attention to aggregated information about the firm's overall performance (instead of disaggregated information about its performance measure in each market or division), such as its product quality ranking and rating of consumer satisfaction. Because of reasons related to coordination, information and influence activities, it is usually difficult to isolate divisions or branches from interventions of the headquarters or influences of other divisions.[23] Moreover, if customers infer the firm's choices from its accounting books, it can prove very difficult to discern accounting records for each of the firm's divisions since there are numerous ways to allocate costs and revenues within the firm.

The second justification, which is related to the first one, is a theoretical one. It is in fact *without loss of generality to use the average signal if the integrated firm can allocate resources and products across markets and divisions freely without being observed by its customers*, e.g., hiring quality control personnel and sending them to individual branches or shipping products from centralized warehouses or production facilities to different markets. Then, even though customers may observe signals from all individual markets, they do not know what is really behind the signal of each market. More formally, suppose that, after $(q_1, q_2, ..., q_n)$ is realized, the integrated firm can choose any profile of qualities $(q_1', q_2', ..., q_n')$ secretly in the n markets as long as $\sum_1^n q_j' = \sum_1^n q_j$ is satisfied. Final signals of individual markets are given by $y_j = q_j' + \theta_j$. In this case, it is intuitively clear that only the sum of qualities or the average quality (hence the average signal) is informative about the big firm's effort choices.

An informal proof goes as follows. To simplify the argument, suppose that there are only two markets and the space of signals is $\Re^2$. The optimal equilibrium must be associated with a subset of signals $\Omega \subset \Re^2$ such that Nash reversion (in every market) occurs with probability 1 if and only if $(y_1, y_2) \in \Omega$ (otherwise $(Buy,\ High)$ will continue to be played and $p = v_h$ is set in

[23] Milgrom and Roberts (1992, p568-576) discuss the advantages and disadvantages of horizontal integration from non-reputation perspectives, and present some interesting case studies such as "IBM and EDS" (page 576) that illustrate the difficulties of maintaining independence for divisions in multidivisional firms.

all the markets), i.e., the optimal equilibrium must have a bang-bang structure. This is because concentrating the harshest punishments (e.g. starting Nash reversion) in the most informative region of signals is more efficient than using weak punishments in less informative regions (cf. Lemma 1). First note that Ω must be symmetric with respect to the $y_1 = y_2$ line. To see this, let Ω' be the mirror image of Ω with respect to $y_1 = y_2$, i.e. $(y_1', y_2') \in \Omega$ if and only if $(y_2', y_1') \in \Omega'$. Then we can construct an equilibrium which uses Ω and Ω' as the punishment region with equal probability and generates the same payoff for the firm. This means that Nash reversion starts with probability $\frac{1}{2}$ when $(y_1, y_2) \in \Omega$ or $(y_1, y_2) \in \Omega'$. If $\Omega \neq \Omega'$, then this means that there exists an optimal equilibrium where Nash reversion occurs with probability less than 1 for some realization of signals. This contradicts the bang-bang property of the optimal equilibrium. Hence Ω and Ω' must coincide, that is, Ω must be symmetric. Next, notice that the the optimal equilibrium remains to be an equilibrium even if the punishment region Ω is replaced by any translate of it to the direction of $(1, -1)$ such as $\Omega + \lambda(1, -1) = \{z \in \Re^2 | y + \lambda(1, -1), y \in \Omega, \lambda \in \Re\}$. This is because these punishment regions are effectively identical to the firm that can reallocate its qualities (q_1, q_2) freely in the direction of $(1, -1)$. Since all these equilibria must be optimal, $\Omega + \lambda(1, -1)$ must be symmetric for all $\lambda \in \Re$ by the same reason as before. For Ω to satisfy this requirement, Ω must look like $\Omega = \{(y_1, y_2) \in \Re^2 | y_1 + y_2 \in K\}$ for some subset K of the real line. Therefore continuation payoffs must depend only on the aggregate signal when characterizing the optimal equilibrium. In this way, we can reduce the dimension of the signals. Note that we do not need the assumption of normal distributions for this proof. We summarize this discussion below as a lemma.

Lemma 3 *Suppose that the firm and the customers repeat the above modified stage game augmented with an additional stage in which the firm can reallocate qualities across markets ex post. Then continuation payoffs must be measurable with respect to the aggregate signal for any optimal equilibrium.*

Given this Lemma, the following analogue of Lemma 1 can be immediately obtained, where the cut-off point for each individual signal is replaced by the cut-off point for the average signal and the consumers from every market behave in a symmetric way.

Proposition 3 *The best equilibrium for the firm is either the repetition of the stage game Nash equilibrium in every market or a cut-off trigger strategy equilibrium in which the following prop-*

erties hold in every market:[24]

(i) In the reputation phase, $(Buy,\ High)$ *is chosen and* p *is set to* v_h *in every market.*

(ii) The play starts in the reputation phase and stay there as long as the average signal y_n *has been no less than a cutoff point* $\widetilde{y}_n$*. Otherwise, the play switches to the punishment phase.*

(iii) In the punishment phase, $(Don't\ Buy,\ Low)$ *is played forever in all the markets.*

Proof: See the Appendix.

Proposition 3 shows that the most effective way to maintain reputation is to use the harshest possible punishment when punishment is called for, which is to punish the firm simultaneously in all markets. Moreover, it needs to depend only on the average/aggregate signal. Proposition 3 also says that in a best equilibrium, the integrated firm should choose high efforts in all n markets in the reputation phase. The idea is that if in an equilibrium customers anticipate that the firm does not choose high efforts in all n markets, they are not willing to pay as high as v_h. Since the firm's profit margin is lower, it has smaller incentives to maintain reputation, thus requiring a higher cut-off point. Since the honesty payoff is lower and the probability of reputation termination is higher, the firm's value per market is lower when it does not choose high efforts in all markets than when it does.

In the following, we focus on such cut-off trigger strategy equilibria.

Characterization

We now characterize the best reputation equilibrium in which the integrated firm chooses high efforts in all n markets. For $j = 0, 1, ..., n$, denote $F_{nj} = F_n(y_n|\bar{e} = e_h - \frac{j}{n}\triangle)$ as the distribution function of y_n when the big firm chooses low efforts in j of n markets. Its average payoff per period in the best reputation equilibrium is given by the following value recursive equation:

$$\Pi = (1-\delta)\, nr + \delta(1 - F_n(\tilde{y}_n|e_h))\Pi \tag{6}$$

where $\tilde{y}_n$ is the cut-off point in the best equilibrium.

For the integrated firm serving n markets, it has n possible deviations by providing low efforts in $m = 1, 2, ..., n$ of the n markets. The IC constraint associated with the mth deviation is

[24] The non-integration case (Lemma 1) is a special case ($n = 1$).

$$\Pi \geq (1-\delta)(nr+md)+\delta(1-F_n(\tilde{y}_n|e_h-\frac{m}{n}\triangle))\Pi \tag{7}$$

To facilitate comparisons, define $\pi_n = \Pi/n$ as the firm's value per market. Then Equations (6) and (7) can be rewritten as

$$\pi_n = (1-\delta)\,r + \delta(1-F_n(\tilde{y}_n|e_h))\pi_n \tag{8}$$

$$\pi_n \geq (1-\delta)(r+\frac{m}{n}d)+\delta(1-F_n(\tilde{y}_n|e_h-\frac{m}{n}\triangle))\pi_n \tag{9}$$

Any pair of $(\pi_n, \tilde{y}_n)$ that satisfies Equation (8) and all the IC constraints of Equation (9) gives rise to a reputation equilibrium in which the big firm chooses high efforts in all markets and customers buy its products as long as the average signal y_n is above $\tilde{y}_n$. As before, the best equilibria feature the smallest cut-off point $\tilde{y}_n$ that satisfies Equation (8) and all the IC constraints of Equation (9). Similar to Lemma 2, it can be shown that one of the IC constraints must be binding at the smallest cut-off point. To solve for the smallest cut-off point $\tilde{y}_n$, we first need to determine which IC constraint is binding.

Suppose the mth IC constraint is binding in the best equilibrium. Parallel to Proposition 1 and Equation (4), the cut-off point in the best equilibrium is the smaller solution to the following fundamental equation:

$$G_{n,m}(y) \equiv \frac{F_n\left(y|e_h-\frac{m}{n}\triangle\right)-F_n\left(y|e_h\right)}{\tau\frac{m}{n}} - F_n\left(y|e_h\right) = \frac{1-\delta}{\delta} \tag{10}$$

The effects of horizontal integration on reputation-building can be clearly seen from Equation (10). Observe that the denominator of the first expression on the LHS is the ratio of the gain from m deviations to the total honesty payoff in n markets. For any fixed m, a larger n means that the punishment for deviations is greater, thus increasing the LHS of Equation (10). This size effect of horizontal integration helps reputation-building by lowering the equilibrium cut-off point $\tilde{y}_n$. On the other hand, a larger n reduces $F_n\left(y|e_h-\frac{m}{n}\triangle\right)-F_n\left(y|e_h\right)$, the numerator of the first expression on the LHS of Equation (10), because deviations in a fixed number of markets are more difficult to detect with a larger n. This will tend to increase the equilibrium cut-off point $\tilde{y}_n$, thus making reputation-building less effective. Moreover, a larger n also means that the merged firm has more sophisticated deviations to contemplate, that is, the number of IC constraints grows with n.

Let's call Equation (9) the $1-$market deviation constraint when $m = 1$.[25] With our information structure $y_n = \bar{e}_n + \bar{\epsilon}_n$, where $\bar{\epsilon}_n$ follows $N(0, \bar{\sigma}_n^2)$, we have the following result.

Proposition 4 *Suppose* $0.5\triangle^2/\sigma_\eta^2 \leq \ln(1+\tau)$. *For any* n, *only the* 1*-market deviation constraint is binding in the best equilibrium for the firm.*

Proof: See the Appendix.

Proposition 4 gives a sufficient condition under which the 1-market deviation IC constraint is the most difficult to satisfy and hence must be binding in the best equilibrium. Roughly speaking, the condition requires that the production noise is significant. In such cases, smaller deviations are much more difficult to detect than larger deviations, thus making the 1-market deviation the most demanding to satisfy.

The condition in Proposition 4 is far from necessary. In particular, normality is not required. Equation (9) is equivalent to

$$(1-\delta)\frac{m}{n}d \leq \delta(F_n(\tilde{y}_n|e_h - \frac{m}{n}\triangle) - F_n(\tilde{y}_n|e_h))\pi_n$$

Note that Equation (9) with $m = 1$ implies all the other constraints for $m > 1$ as long as $F_n(\tilde{y}_n|e_h - \frac{m}{n}\triangle) - F_n(\tilde{y}_n|e_h))$ increases faster than linearly in m. This is satisfied when the density function f_n is increasing around $\tilde{y}_n$.[26] This is usually satisfied at the lower tail of distribution, which is the relevant domain when δ is large. In the rest of the paper, we assume that only the 1-market deviation is binding in the best equilibrium for any n. Then the optimal cut-off point in the best equilibrium $\tilde{y}_n$ is the smaller solution to the following fundamental equation:

$$G_{n,1}(y) \equiv \frac{F_n\left(y|e_h - \frac{1}{n}\triangle\right) - F_n(y|e_h)}{\tau\frac{1}{n}} - F_n(y|e_h) = \frac{1-\delta}{\delta} \tag{11}$$

With the information structure $y_n = \bar{e}_n + \bar{\epsilon}_n$, the fundamental equation of (11) becomes

$$G_n(k) \equiv \frac{\Phi\left(k + \frac{\triangle}{n\bar{\sigma}_n}\right) - \Phi(k)}{\tau\frac{1}{n}} - \Phi(k) = \frac{1-\delta}{\delta} \tag{12}$$

[25] Note that since the firm allocates products across markets evenly, product quality in every market becomes lower if the firm deviates to low effort in one of the n markets.

[26] Convexity of $F_n(\tilde{y}_n|e_h - x\triangle)$ in $x \in [0,1]$ at $\tilde{y}_n$ is sufficient for our purpose. Since $y_n = \bar{e}_n + \bar{\epsilon}_n$, this condition is equivalent to $F_n(\tilde{y}_n + x\triangle|e_h)$ being convex in $x \in [0,1]$, which in turn is equivalent to $f_n(\tilde{y}_n + x\triangle|e_h)$ increasing in $x \in [0,1]$.

where $k = \frac{y - e_h}{\bar{\sigma}_n}$. Note that when $n = 1$, $\bar{\sigma}_1 = \sigma$, and Equation (12) becomes Equation (4). Let $\tilde{k}_n(\delta, \tau, \triangle, \bar{\sigma}_n)$ be the smaller of the two solutions to Equation (12). It is easy to see that $\tilde{k}_n(\delta, \tau, \triangle, \bar{\sigma}_n)$ has all the properties of $\tilde{k}(\delta, \tau, \triangle, \sigma)$, the smaller of the two solutions to Equation (4). That is, $\tilde{k}_n(\delta, \tau, \triangle, \bar{\sigma}_n)$ is increasing in τ and $\bar{\sigma}_n$, and decreasing in δ and $\triangle$.

With this transformation of variables, the probability of reputation termination in the best equilibrium, $F_n(\tilde{y}_n|e_h)$, is simply $\Phi(\tilde{k}_n)$. In summary, the best equilibrium for the firm serving n markets can be characterized as follows.

Proposition 5 *There exists a reputation equilibrium for the firm serving n markets as long as $\delta \geq \delta_n^*$ for some δ_n^*. The optimal cut-off point in the best equilibrium for the firm, $\tilde{k}_n$, is the smaller solution to the fundamental equation (12). The merged firm's value per market is*

$$\pi_n = \frac{(1-\delta)r}{1 - \delta[1 - \Phi(\tilde{k}_n)]}$$

It is increasing in r, δ and $\triangle$, and decreasing in d and $\bar{\sigma}_n$.

From Proposition 5, the construction of the best equilibrium under integration parallels nicely with that under non-integration. We exploit this in the next section to investigate the optimal degree of horizontal integration and to conduct comparative statics.

The Case of Independent Technologies

Even though we think common technology seems to fit reality better, we now briefly consider the case in which the integrated firm adopts independent technologies for all its branches. The case of independent technologies differs from that of common technology in that the noise in the average signal $\bar{\epsilon}_n = \sum_1^n (\eta_j + \theta_j)/n \sim N(0, \bar{\sigma}_n^2)$, where $\bar{\sigma}_n^2 = (\sigma_\eta^2 + \sigma_\theta^2)/n$. Proposition 3 is clearly still valid, so we can focus on the equilibrium in which consumers use the average signal to monitor the integrated firm. By Proposition 4, for independent technologies, as long as $0.5n\triangle^2/\sigma_\eta^2 \leq \ln(1+\tau)$ (i.e., n is not too large), then only the 1-market deviation constraint is binding in the best equilibrium for the firm. It then follows that the normalized signal threshold in the best equilibrium can still be characterized by Equation (12), except that now $\bar{\sigma}_n^2 = (\sigma_\eta^2 + \sigma_\theta^2)/n$, instead of $\sigma_\eta^2 + \sigma_\theta^2/n$. With this modification, Proposition 5 is still valid. It will become clear that all of our analysis and basic results in the next section carry through to the case of independent technologies.

Suppose instead n is sufficiently large so that the condition in Proposition 4 does not hold and the one-market deviation constraint is not the binding constraint. Then it can be shown that the n-market deviation constraint will be binding. Then the normalized signal threshold in the best equilibrium can be characterized by

$$G_n(k) \equiv \frac{\Phi\left(k + \frac{\triangle}{\bar{\sigma}_n}\right) - \Phi(k)}{\tau} - \Phi(k) = \frac{1-\delta}{\delta}$$

The above equation differs from the fundamental equation (4) for the independent firm only in that here the signal noise $\bar{\sigma}_n$ is scaled down by $\sqrt{n}$. By Proposition 2, the profit per market for an integrated firm with independent technologies for its branches will be higher than that of an independent firm. Indeed the complete monopoly is the optimal configuration in this situation and the first best is achieved as $n \to \infty$.

This result is parallel to Matsushima (2001)'s result for multimarket contact, which shows that the first best is approximately achieved by two big firms when the number of the independent markets in which they compete becomes larger. The reason is that the size effect (Bernheim and Whinston, 1990; Andersson, 2002) and the strong information aggregation effect in the case of independent signals together dominate the deviation effect, making integration always better than non-integration. Since we assume explicitly that one source of noise comes from the firm's technology, complete independence of noise would be an extreme assumption to make here. Managing all branches with independent technologies could be prohibitively costly. Furthermore, we do not obtain the interesting result on the bound of firm size when all noises are completely independent. For this reason, we focus on the case where there is *some* common component (see footnote 21) in technology shock across the markets.

5 Optimal Degree of Horizontal Integration

In the two preceding sections we derived the best equilibria under non-integration and integration of n markets (with common technology). By Proposition 5, it is clear that the comparison of non-integration and integration depends on the probability of reputation termination in equilibrium under non-integration, $\Phi(\tilde{k})$, and under integration, $\Phi(\tilde{k}_n)$. Hence, for any given $n > 1$, non-integration dominates integration ($\pi \geq \pi_n$) if and only if $\tilde{k} \leq \tilde{k}_n$.

Since Equation (4) is a special case of Equation (11), a more general question is: what is the optimal degree of horizontal integration? Or, in other words, what is the optimal size of the firm? Conceptually, the answer is straightforward. For all $n = 1, 2, ...$, let n^* be such that $\tilde{k}_n$ is smallest. Then the optimal size of the firm is simply n^*. If $n^* = 1$, then non-integration is optimal. If $n^* > 1$, then a firm serving n^* markets can best maintain reputation.

Proposition 6 *For any δ, $\tilde{k}_n$ is bounded from below and n^* is finite.*

Proof: See the Appendix.

Proposition 6 says that the maximum profit level, r, possible under perfect monitoring, cannot be approximated even if the size of the firm is allowed to go to infinity. It also says that there exists an optimal size of the firm. This is in sharp contrast with Bernheim and Whinston (1990), Matsushima (2001), Andersson (2002), Fishman and Rob (2002), and other papers in the existing literature on reputation, all of which imply that the bigger, the better. Our result differs from these papers because we introduce a common production noise, η, which does not vanish with information aggregation when $n \to \infty$.[27] The optimal size of the firm can be bounded in our model because, as the firm size increases, the positive size and information effects become less and less important, but the negative deviation effect becomes more and more significant since it is more demanding to "detect" one-market deviations for larger firms.

In fact, we can prove a stronger result.

Proposition 7 *Non-integration is optimal when (i) σ_θ is sufficiently small; or (ii) δ is sufficiently close to one; or (iii) τ is sufficiently small.*

Proof: See the Appendix.

Proposition 7 gives several sufficient conditions under which non-integration is optimal ($n^* = 1$). In the first case when the idiosyncratic taste noise is not important, the information aggregation benefit from integration is gone, so non-integration is optimal. In the last two cases, reputation can be maintained quite effectively for firms of all sizes in the sense that the cut-off point can be set at a low level. In such cases, the marginal benefit of the size effect from having more severe punishments is less important. In addition, it is much more difficult to detect small deviations in larger firms in the lower tail of the distribution. As a result, integration brings less benefits but more costs, thus it is dominated by non-integration.

[27] As long as there exists some common noise component, we can allow idiosyncratic components in the production noise as well.

Proposition 8 *The optimal degree of integration n^* is non-increasing in δ.*

Proof: See the Appendix.

Proposition 8 shows that as the discount factor increases, the optimal size of the firm will decrease (at least weakly) and non-integration is more likely to dominate integration. The intuition behind this result is roughly as follows. As δ increases, the future payoffs are more important and hence punishments for deviations are larger. This implies that firms of all sizes can maintain reputation more effectively. That is, the equilibrium cut-off points to continue cooperative actions can be set at low levels. Relatively speaking, the positive size effect of integration is less important in the sense that the marginal benefits of increasing punishments for deviation through integration become smaller. On the other hand, since the equilibrium cut-off points are low, the negative deviation effect of integration becomes more important because low cut-off points make it more difficult to detect a small deviation of a large firm. These forces together imply that as δ increases, the optimal size of the firm will not be larger.

Proposition 9 *The optimal degree of integration n^* is non-decreasing in τ.*

Proof: See the Appendix.

Proposition 9 shows that as τ decreases, the optimal size of the firm will decrease (at least weakly) and non-integration is more likely to dominate integration. Since $\tau = d/r$, it means that a smaller deviation gain, d, or a greater honesty payoff, r, will favor smaller firms and non-integration. The intuition behind this result is similar to that of Proposition 8. A smaller τ means less incentive to deviate and thus smaller or independent firms can build reputation more effectively. Consequently, the marginal benefits of the size effect of integration become less important, while the negative deviation effect of integration is more severe. Therefore, the smaller is τ, the smaller is the optimal size of the firm.

Next we consider how the informativeness of the signal affects the optimal degree of integration. We say that the public signal is uniformly more informative if $\triangle$ is larger (keeping other parameters constant) or if both σ_θ and σ_η are smaller while their ratio is fixed.

Proposition 10 *Suppose δ is sufficiently close to one. The optimal degree of integration, n^*, is non-increasing when the public signal becomes uniformly more informative.*

Proof: See the Appendix.

Proposition 10 shows that for large δ, as the public signal becomes more informative about the firm's effort/quality choices, the optimal size of the firm will decrease (at least weakly) and non-integration is more likely to dominate integration. The intuition behind this result is as follows. When δ is large, the optimal cut-off point can be set quite low. When the public signal becomes more informative, small firms benefit more than larger firms because a small deviation by a larger firm can be "detected" only slightly better with more informative signals. Thus, while more informative signals make firms of all sizes better, the negative deviation effect of integration makes smaller firms benefit more. Therefore, the more informative the public signal, the smaller the optimal size of the firm.

We derive the results of this section under the linear normal information structure. However, from the proofs and the intuition given above, it is not difficult to see that the insights extend more generally. With a general information structure, under reasonable conditions, for smaller y, the n^* that maximizes the function $G_{n,1}(y)$ will be smaller. That is, it is more difficult to "detect" a one market deviation by a larger firm if the cut-off point is in the lower tail of the signal distribution. In such cases, results similar to those obtained here should hold under more general information structure.

6 Conclusion

In this paper, we build a simple model of firm reputation in which customers can only imperfectly monitor firms' effort/quality choices, and then use the model to study the effects of horizontal integration on firm reputation. Our analysis leads to a reputation theory of the optimal size of the firm. Our comparative statics results can be helpful for understanding patterns of horizontal integration in the real world.

This paper has focused on the moral hazard aspects of firm reputation. As is common in this type of model, the firm maintains good reputation on the equilibrium path until a bad realization of the public signal, from which point on the firm enters the punishment phase in which either customers desert the firm or the firm pays large financial penalties. This kind of equilibrium behavior has some unattractive features. First, firm reputation is relatively constant and has no real dynamics. Second, the reversion from good reputation to punishment phases, which is necessary to provide incentives to maintain reputation, depends heavily on coordination of beliefs between the firm and its customers. In equilibrium, punishments are triggered purely by bad luck, not by bad behavior on the firm's part. In addition, when punishments take the form of a permanent end to the relationship, they are not renegotiation-proof.

To deal with some of the above shortcomings, we demonstrate in the Appendix B that there exists an efficient renegotiation-proof equilibrium instead of Nash reversion. However, these issues may be addressed more suitably by introducing adverse selection into the model. Recent contributions by Mailath and Samuelson (2001) and Tadelis (2002) have made important progress in that direction. Introducing adverse selection into our model may not only generate richer reputation dynamics and serve to relax belief coordination requirements, but also may address interesting questions such as: does larger firm size help good-type firms build reputation? Can good-type firms use size to separate themselves from bad types? These questions are left as topics for future research.

In this paper we consider only separate markets and assume away possible linkages across markets, e.g., competition or economies of scales. Those linkages create well understood incentives or disincentives for integration, which we deliberately ignore to focus on how reputation is related to integration. However, reputation and competition may interact and lead to other interesting effects. For example, having a competitor in the market may allow consumers to carry out credible punishment of dishonest behavior by one firm, thus helping the firm build reputation (see, e.g., Hörner, 2002). This would generate a disincentive for the two firms to merge (although merge eliminates price competition and raises joint profit in a static setting). Extending the model in this paper to competitive markets is an interesting topic for future research.

Appendix A: Proofs

Proof of Lemma 1: First suppose that (*Don't Buy*, *Low*) is played in the first period in the best equilibrium. Then it is optimal to play the same equilibrium from the second period on, because the continuation game is isomorphic to the original game. This implies that one possible best equilibrium is to play (*Don't Buy*, *Low*) every period independent of history (with price being set high enough), yielding equilibrium payoffs of $(0,0)$. Note that the repetition of (*Don't Buy*, *Low*) is the only equilibrium outcome which achieves such equilibrium payoffs.

Suppose that the best equilibrium achieves more than $(0,0)$. Neither (*Buy*, *Low*) nor (*Don't Buy*, *High*) can be the first period outcome of the equilibrium which maximizes the firm's payoff. Therefore (*Buy*, *High*) with $p \leq v_h$ should be the outcome of the first period of such equilibrium.

Now we show that such a best equilibrium for the firm must be a *cut-off trigger strategy equilibrium* with $p = v_h$.[28] Let $V^* > 0$ be the best equilibrium payoff for the firm, $p^* (\leq v_h)$ be the equilibrium first period price, and u^* be the mapping which maps each public signal y to the equilibrium continuation payoff $u^*(y) \in [0, V^*]$. Let U be the set of all measurable functions $u : \Re \to [0, V^*]$. Then the following holds:

$$
\begin{aligned}
V^* \quad \leq \quad & \max_{u \in U} (1-\delta)(p^* - c_h) + \delta E[u(y)|e_h] \\
& s.t.\ \ (1-\delta)(p^* - c_h) + \delta E[u(y)|e_h] \geq (1-\delta)(p^* - c_l) + \delta E[u(y)|e_l]
\end{aligned}
$$

where the first inequality comes from the fact that the true set of continuation equilibrium payoffs may not be able to take all the values between 0 and V^*.

However, it is not difficult to show that the (essentially unique) solution $\tilde{u} \in U$ for this optimization problem satisfies $\tilde{u}(y) = 0$ for $y \in (-\infty, \widetilde{y})$ and $\tilde{u}(y) = V^*$ for $y \in [\widetilde{y}, \infty)$ for some $\widetilde{y}$ by the MLRP.[29] Then since both V^* and 0 are equilibrium payoffs, the maximized value of this optimization problem can indeed be achieved as an equilibrium payoff by using the following cut-off trigger strategy (starting in state 1);

- State 1: Play *High* with $p = p^*$ and move to State 2 if and only if $y \in (-\infty, \widetilde{y})$.

[28] The best equilibrium payoff exists because the equilibrium payoff set is compact.

[29] The following perturbation argument might be useful to understand this. Suppose that $u(y') < u(y'')$ for some $y'' < y'$. Consider a perturbation $u(y') + \varepsilon'$ and $u(y'') - \varepsilon''$ for $\varepsilon', \varepsilon'' > 0$ such that $f(y'|e_h)\varepsilon' - f(y''|e_h)\varepsilon'' = 0$. Then the expected continuation payoff given e_h is the same as before, but the expected payoff given e_l is *strictly* lower because $f(y'|e_l)\varepsilon' - f(y''|e_l)\varepsilon'' < 0$ if f satisfies MLRP.

- State 2: Play *Low* with $p > v_l$ and stay at state 2.

Since $\tilde{u} \in U$ is (essentially) the unique solution to the above optimization problem, the equilibrium continuation payoff function u^* to achieve V^* must be $\tilde{u}$ (almost everywhere). Finally, there is no restriction on p^* as long as $p^* \leq v_h$. Thus the optimal price should be $p^* = v_h$ in every period at State 1. *Q.E.D.*

Proof of Proposition 2: It is easy to check that $\lim_{k\to-\infty} G(k) = 0$ and $\lim_{k\to\infty} G(k) = -1$. Furthermore, $G(k)$ is unimodal (pseudoconcave). To see this, note that

$$G'(k) = \frac{\phi(k + \frac{\triangle}{\sigma}) - \phi(k)}{\tau} - \phi(k)$$

It is easy to check that $G'(k^*) = 0$ has a unique solution at

$$k^* = -\frac{\triangle}{2\sigma} - \frac{\sigma ln(1+\tau)}{\triangle}$$

Since the normal distribution satisfies the strict monotone likelihood ration property (Milgrom, 1981; Riley, 1988), it must be that $G'(k) > 0$ for $k \in (-\infty, k^*)$ and $G'(k) < 0$ for $k \in (k^*, \infty)$. Hence k^* maximizes $G(k)$ and $G(k^*) > 0$. Thus, $G(k)$ is unimodal (pseudoconcave).

Since $(1-\delta)/\delta$ is strictly decreasing in δ and goes to zero as δ goes to one, Equation (4) has a solution for all $\delta \geq \delta^*$, where δ^* satisfies $G(k^*) = (1-\delta)/\delta$. For all $\delta > \delta^*$, there are two solutions for Equation (4). By Proposition 1, the cut-off point in the best equilibrium corresponds to the smaller solution for Equation (4), and its equilibrium payoff is given by Equation (5). For $\delta > \delta^*$, all the comparative statics results are verified immediately from Figure 1. *Q.E.D.*

Proof of Proposition 3: The proof is almost identical to the proof of Lemma 1 once we reduce the dimensionality of signals. Now the firm can deviate from e_h to e_l in any subset of the n markets. But the trick to concentrate punishments in the lower tail of the distribution (cf. Footnote 29) still works because MLRP holds with respect to every such deviation in the same direction. Other than this, the proof is completely identical. *Q.E.D.*

Proof of Proposition 4: For any n, let $k = \frac{y-e_h}{\bar{\sigma}_n}$. Let $\widetilde{k}_{nm}$ be the smaller solution to

$$G_{nm}(k) \equiv \frac{\Phi\left(k + \frac{m\triangle}{n\bar{\sigma}_n}\right) - \Phi(k)}{\tau\frac{m}{n}} - \Phi(k) = \frac{1-\delta}{\delta} \tag{13}$$

Just as $G(k)$ in Proposition 2, $G_{nm}(k)$ is unimodal and is maximized at

$$k^*_{nm} = -\frac{1}{2}\frac{m\triangle}{n\bar{\sigma}_n} - \frac{n\bar{\sigma}_n}{m\triangle}\ln(1+\tau\frac{m}{n})$$

Lemma 4 *Suppose* $\ln(1+\tau) \geq 0.5\triangle^2/\sigma^2_\eta$. *Then for any* n *and for any* $m \leq n$, $k^*_{nm} + \frac{m\triangle}{n\bar{\sigma}_n} \leq 0$.

Proof: Let $x = m/n \in (0,1]$. Define $\kappa_{nm}(x)$ as

$$\kappa_{nm}(x) = k^*_{nm} + \frac{m\triangle}{n\bar{\sigma}_n} = \frac{1}{2}\frac{x\triangle}{\bar{\sigma}_n} - \frac{\bar{\sigma}_n}{x\triangle}\ln(1+x\tau)$$

Under the assumption that $\ln(1+\tau) \geq 0.5\triangle^2/\sigma^2_\eta$, we have

$$\kappa_{nm}(x=1) = \frac{1}{2}\frac{\triangle}{\bar{\sigma}_n} - \frac{\bar{\sigma}_n}{\triangle}\ln(1+\tau) < 0$$

Also, as $x \to 0$,

$$\lim_{x\to 0}\kappa_{nm} \to -\lim_{x\to 0}\frac{\tau\bar{\sigma}_n}{\triangle(1+x\tau)} = -\frac{\tau\bar{\sigma}_n}{\triangle} < 0$$

Note that

$$x^2\kappa'_{nm}(x) = \frac{1}{2}\frac{x^2\triangle}{\bar{\sigma}_n} + \frac{\bar{\sigma}_n}{\triangle}\ln(1+x\tau) - \frac{\bar{\sigma}_n}{\triangle}\frac{\tau x}{1+x\tau}$$

Let $\mu(x)$ be the RHS expression. Then $\mu(x=0) = 0$. Moreover,

$$\begin{aligned}
\mu'(x) &= \frac{x\triangle}{\bar{\sigma}_n} + \frac{\tau\bar{\sigma}_n}{\triangle(1+x\tau)} - \frac{\tau\bar{\sigma}_n}{\triangle}\frac{1}{(1+x\tau)^2} \\
&= \frac{x\triangle}{\bar{\sigma}_n} + \frac{\tau\bar{\sigma}_n}{\triangle(1+x\tau)}(1-\frac{1}{1+x\tau}) > 0
\end{aligned}$$

So, $\mu(x) > 0$ and hence $\kappa'_{nm}(x) > 0$ for all $x \in (0,1]$. It follows that $\kappa_{nm}(x) < 0$ for all $x \in (0,1]$. *Q.E.D.*

Since $\tilde{k}_{nm} \le k^*_{nm}$, Lemma 4 implies that for any n and for any $m \le n$, $\tilde{k}_{nm} + \frac{m\triangle}{n\bar{\sigma}_n} \le 0$.

Note that

$$\begin{aligned} \frac{\partial G_{n,m}}{\partial m} &= \frac{\frac{m\triangle}{n\bar{\sigma}_n}\phi\left(k+\frac{m\triangle}{n\bar{\sigma}_n}\right) - \Phi\left(k+\frac{m\triangle}{n\bar{\sigma}_n}\right) + \Phi(k)}{\tau\frac{m^2}{n}} \\ &= \frac{\triangle}{m\tau\bar{\sigma}_n}\left[\phi\left(k+\frac{m\triangle}{n\bar{\sigma}_n}\right) - \phi\left(\bar{k}\right)\right]]]]] \end{aligned}$$

where $\bar{k} \in (k, k+\frac{m\triangle}{n\bar{\sigma}_n})$. Since for any n and for any $m \le n$, $\tilde{k}_{nm}+\frac{m\triangle}{n\bar{\sigma}_n} \le 0$. This implies $\phi\left(k+\frac{m\triangle}{n\bar{\sigma}_n}\right) > \phi\left(\bar{k}\right)$ in the relevant range, and thus $G_{nm}(k)$ is increasing in m. Therefore, only the 1-shot deviation constraint is binding in the best equilibria for any n. *Q.E.D.*

Proof of Proposition 6: We ignore the issue of integer values and treat n as a continuous variable. Since $\bar{\sigma}_n^2 = \sigma_\eta^2 + \sigma_\theta^2/n$, it can be verified that

$$\begin{aligned} \frac{d\frac{1}{n\bar{\sigma}_n}}{dn} &= -\frac{1}{n^2\bar{\sigma}_n}\left(1-\frac{0.5\sigma_\theta^2}{n\bar{\sigma}_n^2}\right) = -\frac{1}{n^2\bar{\sigma}_n} + \frac{0.5\sigma_\theta^2}{n^3\bar{\sigma}_n^3} \\ \frac{d^2\frac{1}{n\bar{\sigma}_n}}{dn^2} &= \frac{2}{n^3\bar{\sigma}_n} - \frac{2\sigma_\theta^2}{n^4\bar{\sigma}_n^3} + \frac{3\sigma_\theta^4}{4n^5\bar{\sigma}_n^5} \end{aligned}$$

When n goes to infinity, we have

$$\begin{aligned} \lim_{n\to\infty} G_n(k) &\longrightarrow \lim_{n\to\infty} \frac{\phi\left(k+\frac{\triangle}{n\bar{\sigma}_n}\right) d\frac{\triangle}{n\bar{\sigma}_n}/dn}{-\tau\frac{1}{n^2}} - \Phi(k) \\ &= \frac{\triangle\phi(k)}{\tau}\lim_{n\to\infty}\left[\frac{1}{\bar{\sigma}_n} - \frac{0.5\sigma_\theta^2}{n\bar{\sigma}_n^3}\right] - \Phi(k) \\ &= \frac{\triangle\phi(k)}{\tau\sigma_\eta} - \Phi(k) \end{aligned}$$

Hence, when n goes to infinity, the fundamental equation (12) becomes

$$\frac{\triangle\phi(k)}{\tau\sigma_\eta} - \Phi(k) = \frac{1-\delta}{\delta}$$

The solution $\tilde{k}$ for this equation (if it exists) is clearly finite. Hence, $\lim_{n\to\infty}\tilde{k}_n > -\infty$.

Next note that $G_n(k) = \frac{\frac{\triangle}{\bar{\sigma}_n}\phi(\widetilde{k})}{\tau} - \Phi(k)$ for some $\widetilde{k} \in \left(k, k + \frac{\triangle}{n\bar{\sigma}_n}\right)$. Remember that relevant k is negative by Lemma 4. For any negative k, there exists n' such that $G_{n'}(k) > \lim_{n\to\infty} G_n(k)$. This implies that n' dominates all large n, which implies that the size of optimal integration is bounded. *Q.E.D.*

Proof of Proposition 7: Differentiating the LHS of Equation (12) with respect to n gives

$$\begin{aligned}\tau\frac{\partial G_n(k)}{\partial n} &= \Phi\left(k + \frac{\triangle}{n\ \bar{\sigma}_n}\right) - \Phi(k) + n\phi\left(k + \frac{\triangle}{n\bar{\sigma}_n}\right)\frac{d\frac{\triangle}{n\bar{\sigma}_n}}{dn} \\ &= \Phi\left(k + \frac{\triangle}{n\ \bar{\sigma}_n}\right) - \Phi(k) - \phi\left(k + \frac{\triangle}{n\bar{\sigma}_n}\right)\frac{\triangle}{n\bar{\sigma}_n}\left(1 - \frac{0.5\sigma_\theta^2}{n\bar{\sigma}_n^2}\right)\end{aligned}$$

This can be rewritten as

$$\frac{\tau n\bar{\sigma}_n^2}{\phi\left(k + \frac{\triangle}{n\bar{\sigma}_n}\right)\frac{\triangle}{n\bar{\sigma}_n}}\frac{\partial G_n(k)}{\partial n} = \triangle\bar{\sigma}_n\frac{\Phi\left(k + \frac{\triangle}{n\ \bar{\sigma}_n}\right) - \Phi(k) - \phi\left(k + \frac{\triangle}{n\bar{\sigma}_n}\right)\frac{\triangle}{n\bar{\sigma}_n}}{\phi\left(k + \frac{\triangle}{n\bar{\sigma}_n}\right)(\frac{\triangle}{n\bar{\sigma}_n})^2} + 0.5\sigma_\theta^2 \tag{14}$$

As $\sigma_\theta \to 0$, the last term goes to zero. Also, $\triangle\bar{\sigma}_n \geq \triangle\sigma_\eta > 0$ for all σ_θ and all n. Let $x = \frac{\triangle}{n\bar{\sigma}_n} \in (0, \frac{\triangle}{\sigma})$. We show that for all x (hence, for all n and all σ_θ), $\exists\xi < 0$ such that

$$\rho(k, x) = \frac{\Phi(k + x) - \Phi(k) - \phi(k + x)\,x}{\phi(k + x)\,x^2} < \xi < 0$$

Note first that $\rho(k, x) < 0$ for all $x > 0$, because the numerator equals $(\phi(\vec{k}) - \phi(k + x))x < 0$, where $\vec{k} \in (k, k + x)$ (assuming that all the cut-off points are small enough). As $x \to 0$, it can be verified that $\rho(k, x) \to k/2 < 0$. Furthermore, one can show that $\partial\rho/\partial x$ has the same sign as

$$-2[\Phi(k + x) - \Phi(k) - \phi(k + x)\,x] + x(k + x)[\Phi(k + x) - \Phi(k)]$$

This function takes a value of zero when $x = 0$ and has a derivative of $(k+x)[\Phi(k + x) - \Phi(k) - \phi(k + x)\,x] + x[\Phi(k + x) - \Phi(k)] > 0$. Thus, $\rho(k, x)$ is increasing in x. Let $\xi = \rho(\frac{\triangle}{\sigma}) < 0$. Then ρ is uniformly bounded above by some $\xi < 0$ (for any small enough k). Therefore, for each fixed k, $\frac{\partial G_n(k)}{\partial n} < 0$ as $\sigma_\theta \to 0$ for all n.

Let k_n^* be the solution of fundamental equation for $G_n(k) = \frac{1-\delta}{\delta}$. If σ_θ is small enough, $G_1(k_1^*) > G_n(k_1^*)$, therefore $k_1^* < k_n^*$ for $n = 2, 3.....$ It follows that non-integration is optimal.

Now consider the case of $k \to -\infty$. Note that as $k \to -\infty$, for all $x \in (0, \frac{\triangle}{\sigma})$,

$$\begin{aligned}
\lim_{k\to\infty}\rho(k,x) &= \lim_{k\to\infty}\frac{\Phi\left(k+x\right)-\Phi\left(k\right)}{\phi\left(k+x\right)x^2}-\frac{1}{x}\\
&= \lim_{k\to\infty}\frac{1-\exp\left(kx+0.5x^2\right)}{-(k+x)x^2}-\frac{1}{x}\\
&= -\frac{1}{x}
\end{aligned}$$

Hence, for any n and for a sufficiently small $\underline{k}_n$, since $x=\frac{\triangle}{n\bar{\sigma}_n}$, the RHS of Equation (14) goes to

$$-\triangle\bar{\sigma}_n\frac{1}{x}+0.5\sigma_\theta^2=-n\bar{\sigma_n}^2+0.5\sigma_\theta^2<0$$

Furthermore, it can be verified that $\partial\rho/\partial k$ has the same sign as

$$(k+x)[\Phi\left(k+x\right)-\Phi\left(k\right)]+\phi\left(k+x\right)-\phi\left(k\right)$$

This function goes to zero as $k\to-\infty$, and has a derivative of $\Phi\left(k+x\right)-\Phi\left(k\right)-x\phi\left(k\right)>0$. Thus, $\rho(k,x)$ is increasing in k.

Take $\underline{k}_1$. We know that the RHS of Equation (14) is negative at $x=\frac{\triangle}{\bar{\sigma}}$ as long as $k\leq\underline{k}_1$. Since ρ is increasing in x, the RHS of Equation (14) must be negative for all $x<\frac{\triangle}{\bar{\sigma}}$ when $k=\underline{k}_1$. Moreover, since ρ is increasing in k, this must be true for all x as long as $k\leq\underline{k}_1$.

Therefore, for sufficiently small k, it must be that G_n is decreasing in n so that $G_1(k)>G_n(k)$ for all $n>1$. Since $\tilde{k}_n$ goes to $-\infty$ for all n when either δ is sufficiently close to one or τ is sufficiently small, Non-integration is optimal ($n^*=1$) as before in either of the cases. *Q.E.D.*

Proof of Proposition 8: We first prove a useful lemma. Define $G_0(k;t,x)=\frac{\Phi(k+x)-\Phi(k)}{t}-\Phi(k)$.

Lemma 5 *Let $0<t'<t$ and $0<x'<x$. There exists a $\hat{k}$ such that $G_0(k;t',x')\leq G_0(k;t,x)$ if and only if $k\leq\hat{k}$.*

Proof of Lemma 5: Define $w(k;t,x,t',x')=G_0(k;t,x)-G_0(k;t',x')$. Note that as $k\to-\infty$, both $G_0(k;t,x)$ and $G(k;t',x')$ go to zero; and as $k\to\infty$, both $G(k;t,x)$ and $G(k;t',x')$ go to -1. Thus, $w(k)$ goes to zero as $k\to-\infty$ and as $k\to\infty$.

Differentiating w gives

$$
\begin{aligned}
w'(k) &= \frac{\phi(k+x)-\phi(k)(1+t)}{t} - \frac{\phi(k+x')-\phi(k)(1+t')}{t'} \\
&= \frac{\phi(k)}{t}[e^{-\frac{x^2}{2}}e^{-kx} - 1 - \frac{t}{t'}e^{-\frac{(x')^2}{2}}e^{-kx'} + \frac{t}{t'}]
\end{aligned}
$$

Let us call the expression in the bracket above $\eta(k)$. Since $x > x'$, $\eta(k) \to \infty$ as $k \to -\infty$. When $k \to \infty$, $\eta(k) \to t/t' - 1 > 0$. Furthermore, we have

$$
\eta'(k) = -xe^{-\frac{x^2}{2}}e^{-kx} + x'\frac{t}{t'}e^{-\frac{(x')^2}{2}}e^{-kx'} = e^{-kx}\left[-xe^{-\frac{x^2}{2}} + x'\frac{t}{t'}e^{-\frac{(x')^2}{2}}e^{k(x-x')}\right]
$$

Setting $\eta'(k) = 0$ we obtain

$$
\bar{k} = \frac{ln(xt') - ln(x't) - 0.5x^2 + 0.5(x')^2}{x - x'}
$$

It is clear that $\eta'(k) < 0$ for $k < \bar{k}$ and $\eta'(k) > 0$ for $k > \bar{k}$. So $\eta(k)$ reaches its minimum at $\bar{k}$. If $\eta(\bar{k}) \geq 0$, then $\eta(k) > 0$ for all $k \neq \bar{k}$. This implies that $w'(k) = \phi(k)\eta(k)/t$, $w'(k) > 0$ for all $k \neq \bar{k}$ and $w'(k) = 0$ at $k = \bar{k}$. Thus $w(k)$ is always strictly increasing except at $\bar{k}$ as a reflection point. But this contradicts the fact that $w(k)$ goes to zero as $k \to -\infty$ and as $k \to \infty$. Therefore, it must be that $\eta(\bar{k}) < 0$.

Since $\eta(-\infty) = \infty$ and $\eta(\infty) = t/t' - 1 > 0$, there exist k_1 and k_2, $k_1 < \bar{k} < k_2$, such that $\eta(k) = 0$. So $\eta(k) > 0$ for $k \in (-\infty, k_1) \cup (k_2, \infty)$ and $\eta(k) < 0$ for $k \in (k_1, k_2)$. Since the sign of $w'(k)$ is identical to that of $\eta(k)$, $w(k)$ is strictly increasing for $k \in (-\infty, k_1) \cup (k_2, \infty)$ and strictly decreasing $k \in (k_1, k_2)$. Therefore, there must exist a unique $\hat{k} \in (k_1, k_2)$ such that $w(k) = 0$. Moreover, $w(k) > 0$ for $k < \hat{k}$ and $w(k) < 0$ for $k > \hat{k}$. *Q.E.D.*

Consider any $n < n'$. Let $t = \tau/n$, $x = \triangle/(n\bar{\sigma}_n)$, and $t' = \tau/n'$, $x' = \triangle/(n'\bar{\sigma}_{n'})$. Then, $t > t'$ and $x > x'$. Since $G_n(k) = G_0(k; t, x)$ and $G_{n'}(k) = G_0(k; t', x')$, by Lemma 5, there exists a $\hat{k}_{nn'}$ such that $G_n(k) > G_{n'}(k)$ if and only if $k < \hat{k}_{nn'}$.

Now suppose for some δ, the optimal degree of integration is $n^*(\delta)$. This means that at $\tilde{k}_{n^*(\delta)}$, $G_{n^*(\delta)} \geq G_{n'}$ for all n'. Therefore, for $n' > n^*(\delta)$, it must be that $\tilde{k}_{n^*(\delta)} < \hat{k}_{n^*(\delta)n'}$; and for $n' < n^*(\delta)$, it must be that $\tilde{k}_{n^*(\delta)} > \hat{k}_{n^*(\delta)n'}$.

Consider an increase in δ to $\delta_1 > \delta$. For any n, $\tilde{k}_n$ decreases in δ. In particular, $\tilde{k}_{n^*(\delta)}(\delta_1)$ is smaller than $\tilde{k}_{n^*(\delta)}(\delta)$. Note that δ does not effect $\hat{k}_{nn'}$ at all. Hence, for $n' > n^*(\delta)$, we have $\tilde{k}_{n^*(\delta)}(\delta_1) < \hat{k}_{n^*(\delta)n'}$, so $G_{n^*(\delta)} \geq G_{n'}$ at $\tilde{k}_{n^*(\delta)}(\delta_1)$. Therefore, the optimal degree of integration at δ_1, $n^*(\delta_1)$, cannot be greater than $n^*(\delta)$. For all $n' < n^*(\delta)$, if $\tilde{k}_{n^*(\delta_1)} > \hat{k}_{n^*(\delta)n'}$, then we must have $n^*(\delta_1) = n^*(\delta)$. Otherwise, if $\tilde{k}_{n^*(\delta_1)} < \hat{k}_{n^*(\delta)n'}$ for some $n' < n^*(\delta)$, then $G_{n^*(\delta)} < G_{n'}$ at $\tilde{k}_{n^*(\delta_1)}$, which implies that $n^*(\delta_1)$ must be smaller than $n^*(\delta)$. *Q.E.D.*

Proof of Proposition 9: The proof is similar to that of Proposition 8. Suppose for some τ, the optimal degree of integration is $n^*(\tau)$. Consider a decrease in τ to $\tau_1 < \tau$. Since G_n is decreasing in τ, then $\tilde{k}_n$ is increasing in τ. Hence, $\tilde{k}_{n^*(\tau)}(\tau_1)$ is smaller than $\tilde{k}_{n^*(\tau)}(\tau)$. From the proof of Lemma 5, for any n and n', $\hat{k}_{nn'}$ depends on the ratio of t/t' but not on t nor t'. Since $t/t' = n/n'$ is independent of τ, for any n and n', $\hat{k}_{nn'}$ is independent of τ. That is, the relative positions of G_n are independent of τ. Therefore, a decrease in τ is like an increase in δ. The same argument in the proof of Proposition 8 applies. *Q.E.D.*

Proof of Proposition 10: We only need to consider an increase in $\triangle$. The argument is identical for a decrease in σ_θ and σ_η of the same proportions. Suppose for some $\triangle$, the optimal degree of integration is $n^*(\triangle)$. Consider an increase in $\triangle$ to $\triangle_1 > \triangle$. Since G_n is increasing in $\triangle$, $\tilde{k}_n$ is decreasing in $\triangle$. Hence, $\tilde{k}_{n^*(\triangle)}(\triangle_1)$ is smaller than $\tilde{k}_{n^*(\triangle)}(\triangle)$. From the proofs of Propositions 8 and 9, it suffices to show that for any n and $n' > n$, $\hat{k}_{nn'}$ is nondecreasing in $\triangle$.

Since $G_n = G_{n'}$ at $\hat{k}_{nn'}$, $\hat{k}_{nn'}$ is the solution to $w(k) = 0$, where

$$w(k) = n\left[\Phi\left(k + \frac{\triangle}{n\ \bar{\sigma}_n}\right) - \Phi(k)\right] - n'\left[\Phi\left(k + \frac{\triangle}{n'\ \bar{\sigma}'_n}\right) - \Phi(k)\right]$$

Exactly as in Lemma 5 and Propositions 8, w crosses zero only once at $\hat{k}_{nn'}$ as it is decreasing. So to show $\hat{k}_{nn'}$ is nondecreasing in $\triangle$, we only need to show that $w(k)$ is increasing in $\triangle$ at $\hat{k}_{nn'}$, or $\partial w/\partial\triangle > 0$ at $\hat{k}_{nn'}$.

Since

$$\frac{\partial w}{\partial \triangle} = \frac{1}{\bar{\sigma}_n}\phi\left(k + \frac{\triangle}{n\ \bar{\sigma}_n}\right) - \frac{1}{\bar{\sigma}'_n}\phi\left(k + \frac{\triangle}{n'\ \bar{\sigma}'_n}\right)$$

$\partial w/\partial\triangle > 0$ if $R(n) = \frac{1}{\bar{\sigma}_n}\phi\left(k + \frac{\triangle}{n\ \bar{\sigma}_n}\right)$ is decreasing at $\hat{k}_{nn'}$. It can be verified that

$$\frac{\partial R}{\partial n} = \frac{\phi\left(k + \frac{\triangle}{n\ \bar{\sigma}_n}\right)}{n^2\bar{\sigma}_n^3}\left[\triangle\left(k + \frac{\triangle}{n\ \bar{\sigma}_n}\right)\left(\frac{\sigma_\eta^2}{\bar{\sigma}_n} + \frac{0.5\sigma_\theta^2}{n\bar{\sigma}_n}\right) + 0.5\sigma_\theta^2\right]$$

It is easy to see that $\frac{\partial R}{\partial n} < 0$ for all n if $k < -\frac{0.5\sigma_\theta^2}{\sigma_\eta \triangle} - \frac{\triangle}{\sigma}$. If $\hat{k}_{nn'} < -\frac{0.5\sigma_\theta^2}{\sigma_\eta \triangle} - \frac{\triangle}{\sigma}$, then the proposition holds. Suppose $\hat{k}_{nn'} \geq -\frac{0.5\sigma_\theta^2}{\sigma_\eta \triangle} - \frac{\triangle}{\sigma}$. For large δ, $\tilde{k}_{n^*(\triangle)}(\triangle_1)$ and $\tilde{k}_{n^*(\triangle)}(\triangle)$ are smaller than $-\frac{0.5\sigma_\theta^2}{\sigma_\eta \triangle} - \frac{\triangle}{\sigma}$. Since $\hat{k}_{nn'}$ is independent of δ, then for large δ, $\hat{k}_{nn'}$ is greater than $\tilde{k}_{n^*(\triangle)}(\triangle_1)$ and $\tilde{k}_{n^*(\triangle)}(\triangle)$, so the claim of the proposition is true. *Q.E.D.*

Appendix B: Renegotiation-Proof Equilibrium

In this Appendix, we show that the same outcome derived in Section 2 can be implemented in a renegotiation-proof equilibrium. Consider the following strategy. If the public signal falls below the cut-off point $\tilde{k}$, in the next period the firm offers customers a (very low) price p' and continues to provide high effort, and customers continue to buy from the firm. If the public signal in the punishment period is above another cut-off point k', then the firm is "redeemed," and can switch back to the reputation phase, charging $p = v_h$ in the next period. Otherwise, the firm stays in the punishment phase offering the low price p'.[30] Since the firm provides high effort in every period in both reputation and punishment phases, any such reputation equilibrium is efficient.

For this punishment scheme together with the reputation phase described above to constitute an equilibrium, we need to have

$$(1-\delta)(p'-c_h) + \delta\,(1-\Phi(k'))\,\pi = 0 \qquad (15)$$

where π is the firm's average value as characterized in Proposition 1. The first term, $(1-\delta)(p'-c_h)$, is the firm's (negative) profit per period in the punishment phase averaged over periods; while the second term is the discounted expected future profit if it redeems itself, which occurs with probability $1-\Phi(k')$. Therefore this condition states that p' and k' should be chosen so that the firm's expected payoff is zero once it is in the punishment phase.

In addition, the firm must be willing to provide high effort in the punishment phase, which requires the following incentive constraint:

$$(1-\delta)(p'-c_h) + \delta\,(1-\Phi(k'))\,\pi \geq (1-\delta)(p'-c_l) + \delta\left(1-\Phi(k'+\frac{\triangle}{\sigma})\right)\pi$$

[30] This is similar to a "stick and carrot" equilibrium (Abreu, 1988).

Or, recalling that $d = c_h - c_l$, we have

$$(1-\delta)d \leq \delta\left(\Phi(k' + \frac{\triangle}{\sigma}) - \Phi(k')\right)\pi \tag{16}$$

This simply says that the gain from a one period deviation to low effort during the punishment phase is less than the loss of future profit from low effort which serves to reduce the probability of switching back to the high price and high profit of the reputation phase.

Note that the IC constraint of equation (16) is identical to that of the reputation phase, Equation (2) or (3). Thus, (16) is satisfied if and only if the cut-off point k' is not less than the smaller solution and not greater than the larger solution to the fundamental equation (4). For any such k', if there exists a price p' satisfying equation (15), then we have an efficient renegotiation-proof equilibrium

Two remarks are in order here. First, there may be multiple (k', p') that satisfy the above two conditions (15) and (16). Clearly, the larger k' is, the higher p' is. Second, to satisfy equation (15) may require a negative price p', which may not be feasible in many contexts. By Equation (15), p' can be made as high as possible when k' is the largest among those which satisfy (16). The largest such k' is $k' = -\left(\widehat{k} + \frac{\triangle}{\sigma}\right)$, where $\widehat{k}$ is the cut-off point in the reputation phase, due to the symmetry of normal distribution. Thus, there exists $(k', p'(\geq 0))$ that satisfy (15) and (16) if and only if

$$-(1-\delta)c_h + \delta\Phi(\widehat{k} + \frac{\triangle}{\sigma})\pi \leq 0 \tag{17}$$

This inequality is satisfied when δ is large enough. To summarize, we have the following result.

Proposition 11 *There exist $k' \in \Re$ and $p' \in \Re^+$ that satisfy (15) and (16) if and only if (17) is satisfied. Moreover, (17) is satisfied when δ is sufficiently large.*[31]

Proof: We only show that (17) is satisfied when δ is sufficiently large. (17) is equal to

$$\frac{\delta\Phi(\widehat{k} + \frac{\triangle}{\sigma})\pi}{1-\delta} \leq c_h$$

First we can show that $\frac{\Phi(\widehat{k}+\frac{\triangle}{\sigma})}{1-\delta} \to \tau$ as $\delta \to 1$ as follows:

$$\lim_{\delta\to1} \frac{\Phi(\widehat{k} + \frac{\triangle}{\sigma})}{1-\delta} = \lim_{\delta\to1} -\phi(\widehat{k} + \frac{\triangle}{\sigma}) \cdot \frac{\partial\widehat{k}}{\partial\delta}$$

[31] It can also be shown that if this "stick and carrot" strategy cannot implement the minmax payoff 0 for the firm, then there is no efficient renegotiation-proof equilibrium that can do so.

$$= \lim_{\delta \to 1} \frac{\phi(\widehat{k} + \frac{\triangle}{\sigma})/\delta^2}{\frac{\phi(\widehat{k}+\frac{\triangle}{\sigma})-\phi(\widehat{k})}{\tau} - \phi(\widehat{k})} \text{ (by Equation 4)}$$

$$= \lim_{\delta \to 1} \frac{\phi(\widehat{k} + \frac{\triangle}{\sigma})/\phi(\widehat{k})}{\frac{\phi(\widehat{k}+\frac{\triangle}{\sigma})/\phi(\widehat{k})-1}{\tau} - 1}$$

Since $\widehat{k} \to -\infty$, thus $\phi(\widehat{k} + \frac{\triangle}{\sigma})/\phi(\widehat{k}) \to \infty$ as $\delta \to 1$, we have

$$\lim_{\delta \to 1} \frac{\Phi(\widehat{k} + \frac{\triangle}{\sigma})}{1-\delta} \to \tau$$

Then, since $\pi \to r$ as $\delta \to 1$, we have

$$\lim_{\delta \to 1} \frac{\delta \Phi(\widehat{k} + \frac{\triangle}{\sigma})\pi}{1-\delta} = d = c_h - c_l < c_h$$

Q.E.D.

This renegotiation-proof best reputation equilibrium exhibits a particular kind of price dynamic. The price dynamics of a best reputation equilibrium characterized in this section is shown in Figure 2 below. An example of such price dynamics is an airline company who just had a bad incident (e.g., a plane crash). Even if the incident can be purely bad luck, the company typically offers large discounts to "win back" customers; and such discounts are phased out over time as customers "regain" confidence in the company.

This is similar to the price dynamics of Green and Porter (1984), the first to construct public strategy equilibria with punishment phases in a model of imperfect monitoring. In their equilibrium construction of a repeated duopoly model with stochastic demand, firms continue to collude until the price drops below a threshold level, then they play the Nash Cournot equilibrium for a fixed number of periods before reverting back to the collusive phase. Our construction of the best equilibrium with efficient punishment phases differs from theirs because the firm and customers in our model can use prices to transfer utilities, achieving the efficient outcome at every history.

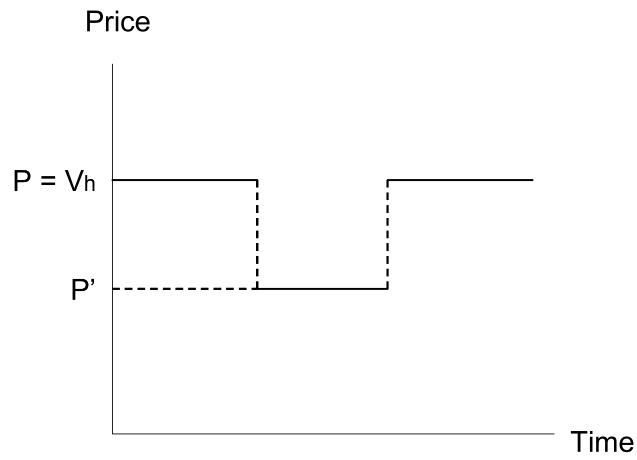

Figure 2:

References

[1] Abreu, Dilip, Paul Milgrom, and David Pearce, 1991, Information and timing in repeated partnerships, *Econometrica*, 59:1713-1733.

[2] Abreu, Dilip, David Pearce, and Ennio Stacchetti, 1986, Optimal Cartel Equilibria with Imperfect Monitoring, *Journal of Economic Theory*, 39:251-269.

[3] Abreu, Dilip, David Pearce, and Ennio Stacchetti, 1990, Towards a Theory of Discounted Repeated Games with Imperfect Monitoring, *Econometrica*, 58:1041-1063.

[4] Alchian, Armen and Harold Demsetz, 1972, Production, Information Costs, and Economic Organization, *American Economic Review*, 62:777-795.

[5] Andersson, Fredrik, 2002, Pooling Reputations, *International Journal of Industrial Organization*, 20:715-730.

[6] Bernheim, Douglas and Michael Whinston, 1990, Multimarket Contact and Collusive Behavior, *RAND Journal of Economics* , 21:1-26.

[7] Cabral, Luis, 2000, Stretching Firm and Brand Reputation, *RAND Journal of Economics*, 31:658-673.

[8] Cabral, Luis, 2007, Optimal Brand Umbrella Size, *Working Paper, New York University*.

[9] Cai, Hongbin and Ichiro Obara, 2004, Firm Reputation and Horizontal Integration, mimeo, UCLA.

[10] Coase, Ronald, 1937, The Nature of the Firm, *Economica*, 4:386-405.

[11] Fishman, Arthur, 2005, Financial Intermediaries as Facilitators of Reputation Acquisition in Credit Market, Working Paper, Bar Ilan University.

[12] Fishman, Arthur and Rafael Rob, 2002, Is Bigger Better? Investing in Reputation, Working Paper, University of Pennsylvania.

[13] Fudenberg, Drew, David Levine and Eric Maskin, 1994, The Folk Theorem with Imperfect Public Information, *Econometrica*, 62:997-1039.

[14] Green, Edward and Robert Porter, 1984, Noncooperative Collusion under Imperfect Price Information, *Econometrica*, 52: 87-100.

[15] Gutman, Joel and Nira Yekouel, 2002, Economics of Scale in Reputation: A New Theory of the Emergence of Firms, Working Paper, Bar Ilan University.

[16] Hart, Oliver, 1995, *Firms, Contracts, and Financial Structure*, Oxford University Press.

[17] Holmstrom, Bengt and John Roberts, 1998, The Boundaries of the Firm Revisited, *Journal of Economic Perspective*,12:73-94.

[18] Hörner, Johannes, 2002, Reputation and Competition, *American Economic Review*, 92:644-663.

[19] Jin, Ginger and Phillip Leslie, 2004, Reputational Incentives for Restaurant Hygiene, Working Paper, Stanford University.

[20] Klein, Benjamin and Keith Leffler, 1981, The Role of Market Forces in Assuring Contractual Performance, *Journal of Political Economy*, 89:615-641.

[21] Kreps, David, 1990, Corporate Culture and Economic Theory, in *Perspectives on Positive Political Economy*, eds. James E. Alt and Kenneth A. Shepsle, Cambridge University Press: New York.

[22] Mailath, George and Larry Samuelson, 2001, Who Wants a Good Reputation? *Review of Economic Studies*, 68:415-441.

[23] Marvel, Howard and Lixin Ye, 2004, Trademark Sales, Entry, and the Value of Reputation, Working Paper, Ohio State University.

[24] Matsushima, Hitoshi, 2001, Multimarket Contact, Imperfect Monitoring, and Implicit Collusion, *Journal of Economic Theory*, 98:158-178.

[25] Milgrom, Paul, 1981, Good News and Bad News: Representation Theorems and Applications, *Bell Journal of Economics*, 12:380-391.

[26] Milgrom, Paul and John Roberts, 1992, *Economics, Organization, and Management*, Englewood Cliffs, NJ: Prentice Hall.

[27] Riley, John, 1988, Ex Post Information in Auctions, *Review of Economic Studies*, 55:409-430.

[28] Rob, Rafael and Tadashi Sekiguchi, 2004, Reputation and Turnover, Working Paper, University of Pennsylvania.

[29] Tadelis, Steven, 1999, What's in a Name? Reputation as a Tradeable Asset, *American Economic Review*, 89:548-563.

[30] Tadelis, Steven, 2002, The Market for Reputation as an Incentive Mechanism, *Journal of Political Economy*, 110:854-882.

[31] Williamson, Oliver, 1985, *The Economic Institutions of Capitalism*, New York: Free Press.